Pan-Canadian Genome Library
Bibliothèque génomique pancanadienne

PAN-CANADIAN GENOME LIBRARY (PCGL)

DATA GOVERNANCE FRAMEWORK

No	Version	Date	Description
1	v.1.0	March 2025	-First internal draft
2	v. 2.0	July 2025	-Revised Final Version
3	V 2.0w	December 2025	Final Version adapted for Website

TABLE OF CONTENTS

Acronyms

DAC: Data Access Committee

REB: Research Ethics Board

PCGL: Pan-Canadian Genome Library

WG: Working Group

Definitions

Access: To retrieve or copy a digital, conceptual or physical asset (including a dataset), in whole or in part.

Aggregate Data: Collated results of data from many individuals included within a specific dataset in PCGL.

Approved Researcher: A researcher who has been approved by the PCGL Data Access Committee (DAC) to access specific datasets.

Controlled-access Data: Data that is available and accessible subject to conditions and a PCGL approval process.

Data Access Agreement: An agreement between PCGL and the recipient of data that describes conditions of transfer, use and disclosure.

Dataset: A collection of data which is part of PCGL.

Discovery: To search for and thereby determine the existence of a digital, conceptual or physical asset (including a dataset), in whole or in part.

Federated Data: Data from Submitters, which cannot be transferred to PCGL for centralized storage, due, for instance, to institutional, consent, regulatory or other limitations.

Metadata: Data about stored data including its properties, history and versions.

Open-access Data: Data that is available and accessible without restrictions.

Prospective Collection: Data from research participants collected and consented specifically for inclusion in PCGL.

Pan-Canadian Genome Library (PCGL): Refers to PCGL's core infrastructure, including, but not limited to, the collection of data and information that is managed and stored in a systematic way to enable data access and discovery.

Research Portal: Online portal that serves as the user interface facilitating both data discovery and access to PCGL data.

Retrospective Collection: Data that has been collected or is being collected from research participants under an REB approval obtained prior to the creation of PCGL (to integrate data into PCGL, the retrospective collection needs to have obtained local approval to do so).

Submitter: Individual researcher, project team or consortium at a participating Canadian institution who wishes to submit to PCGL their participants' clinical, phenotypic and molecular data.

1. Overview

1.1. Background

Canada is in an ideal position to capitalize on the opportunities afforded by genomic data – it boasts world-leading expertise in genome science research and several large-scale genomic projects have been funded and are collecting data on an ongoing basis. Indeed, the size and complexity of human genomics datasets and their associated clinical data are growing rapidly.

However, Canada lacks a formal national, pan-Canadian, strategy on the capture, storage and access to Canadian data in a secure, accessible, and economical manner. Currently, several genomic projects are operating either independently or as part of disease-specific networks. The Pan-Canadian Genome Library's objective is to unify Canada's genome sequencing efforts by creating a federated data management system grouping available datasets and making them available to Approved Researchers.

1.2. Objective of the PCGL project

PCGL proposes a common platform for the storage and sharing of Canadian genomic datasets, both from Prospective and Retrospective Collections, meeting the requirements outlined in this framework.

More specifically, PCGL will provide a common database environment where a Submitter's genomic and phenotypic data will be made available for data discovery and access through a Research Portal**.** PCGL's architecture will encourage integrated (or 'centralized') datasets for efficiency, scalability, and sustainability; and also enable federated datasets for interoperability and flexibility (e.g., for Indigenous datasets).

Given that PCGL proposes a pan-Canadian repository, with datasets from different projects, disease areas, patient populations and communities, particular attention will be paid to adopting a governance model that is appropriate and representative of the diverse data provenance backgrounds. In particular, a Knowledge Users Committee, Patient-partner team, a Data Diversity & Inclusion Working Group and an Indigenous Genetics Circle are involved in advising the development and operations of PCGL.

1.3. Purpose of this Governance Framework

This governance framework is a public document that has been developed to facilitate the early adoption of an overall governance structure for the data managed as part of the PCGL project.

The framework accounts for different data collection sectors and Canadian jurisdictions involved, while aligning with similar efforts globally. It aims to provide useful guidance regarding the administration, custodianship and sharing of data available through PCGL. In addition, it sets out the core requirements for the inclusion of data from Canadian research collections (prospective and retrospective). Finally, this Framework sets out ethical considerations pertaining to the deposit and storage of data in PCGL.

The development and implementation of an ethically and legally robust governance structure is a prerequisite to the creation of a data platform such as PCGL, as it provides immediate guidance and encourages transparency for the involved institutions, researchers, participants/patients, communities and other stakeholders involved in contributing to the development of the platform or in using PCGL data. In practice, PCGL will promote responsible open science by following FAIR data standards and relevant policies developed by the Global Alliance for Genomics and Health (GA4GH), in conformity with applicable Canadian laws and research ethics.

This framework does not directly apply to access to Indigenous data. Policies and procedures related to such access will be developed by the Indigenous Genetics Circle. PCGL is committed to respecting the leadership of Indigenous communities and we will carefully follow the principles and practices they have highlighted and pioneered, in particular the sovereign-based governance models. However, the Indigenous Genetics Circle may still find this governance framework a useful source of information that will facilitate interoperability.

Furthermore, PCGL is committed to promoting diversity and inclusion not just in data gathering, but also in how data is processed, interpreted, shared, and communicated. This work will be spearheaded by the Data Diversity and Inclusion Working Group.

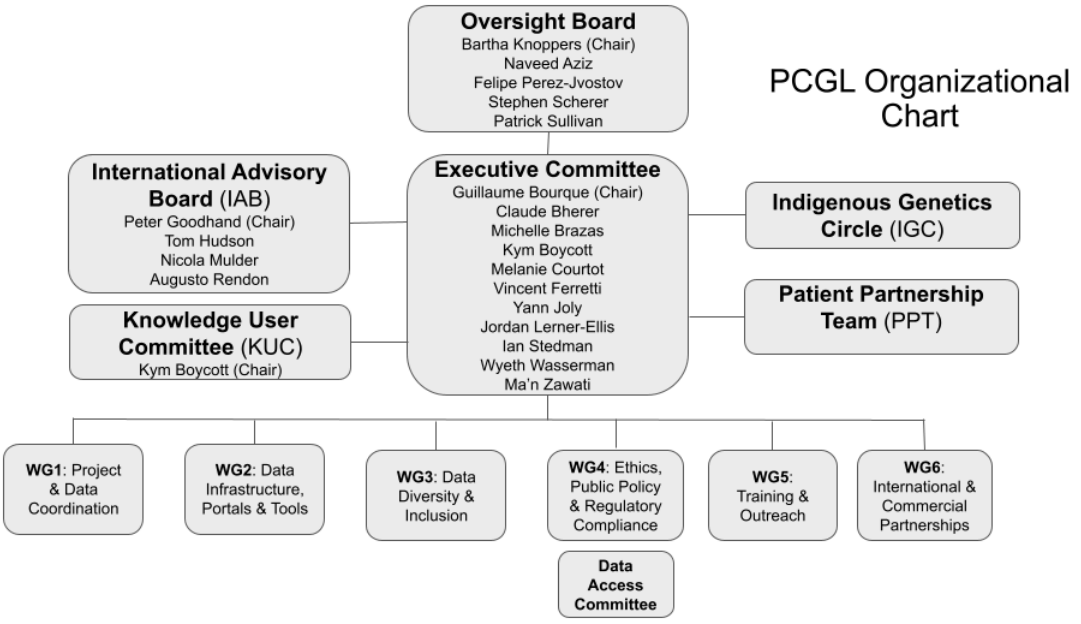
This governance framework will be updated on an as-needed basis and, at minimum, will undergo a formal review and revision at least once annually. Any new iteration that comports substantial changes will be approved by the McGill Research Ethics Board.

2. Governance and organizational structure

2.1. Governance overview

The organizational chart below outlines the various bodies tasked with overseeing, advising, and leading the scientific, administrative, and operational development of PCGL.

Figure 1: PCGL Data Organizational Chart



2.2. Scientific and administrative oversight

PCGL is an inter-institutional project currently under the custodianship of McGill University. In time, PCGL may evolve toward a different governance model, which could bring changes to how it is managed and institutionally supported.

The following section provides an overview of the PCGL project's governance, with a particular emphasis on roles and responsibilities related to data governance.

Oversight Board:

- The Oversight Board (OB) is composed of PCGL stakeholder representatives, including McGill University (as the lead institution, responsible for PCGL's data management and security), Canada's national platform for sequencing and analysis (CGEn) (national-scale genomic data generation and stewardship, including HostSeq data), the Digital Research Alliance of Canada (DRAC), as well as funders, a knowledge user, patient partner and Indigenous representative (TBC). The Board meets quarterly for the first year and biannually after that, with one in-person meeting per year. The Board has the following responsibilities: act to endorse the major decisions, policies, priorities and strategic plans proposed by the Executive Committee; provide guidance on PCGL deliverables; and monitor progress towards PCGL objectives.

Executive Committee:

- The Executive Committee (EC) oversees all aspects of the PCGL project - scientific vision, coordination between WGs (supplemented by having several team members associated with multiple WGs), strategic decision-making to optimize scientific outcomes and benefits to Canadians, and procuring complementary funding as needed to support ancillary elements of the PCGL. The EC meets quarterly, with at least one in-person meeting per year. The EC is accountable to the OB.

International Advisory Board:

- The International Advisory Board (IAB) includes an independent group of distinguished and diverse experts who will support strategic planning by providing global insights to the EC, including research progress, potential international collaborations and emerging technologies. The IAB meets in person annually, with interim updates via teleconference. IAB meetings will be scheduled in advance of OB meetings and the IAB will produce a meeting report to be shared with the OB.

Knowledge Users Committee:

- The Knowledge Users Committee (KUC) consists of individuals who are likely to use the PCGL and make informed decisions about health policies, programs and/or practices. For this framework, users could be those contributing/storing data and those using the data in the PCGL. The chair is part of the EC and will share the progress of the overall project with the KUC, including progress towards depositing initial collections into the PCGL. The chair will also bring KUC perspectives to the attention of the EC for their consideration. This bidirectional flow of information will ensure that KUs have the earliest opportunity to provide feedback on proposed library processes and requirements, for example: data ingestion and archiving, sharing, linkage, analysis. This committee meets virtually on a bi-annual basis.

Indigenous Genetics Circle:

- The IGC has been created to broaden Indigenous expertise to support engagement and consultation across the WGs. It advises on the process to access Indigenous datasets/collections.

Patient Partnership Team:

- The Patient Partnership Team (PPT) works closely with other members of the PCGL WGs to ensure that patients are consulted and included in decision making as the Library is developed.

Working Groups:

- Each WG has a clear mandate together with an identified lead (or co-leads). These include: WG1 Project and data coordination, WG2 Data infrastructure, portals and tools, WG3 Data diversity and inclusion, WG4 Ethics, public policy and regulatory compliance, WG5, training & outreach and WG6 International and commercial partnerships. Each WG meets periodically, and the lead will ensure that progress is made towards the WG's aims and deliverables.

Data Access Committee:

- The DAC is a committee composed of experts responsible for reviewing data access requests submitted to PCGL and authorizing the use of controlled-access data. The committee's composition and the procedures it follows are described in Appendix A.

2.3. Ethical oversight

Ethics approval has been obtained from McGill University's Faculty of Medicine's Research Ethics Board (REB), which serves as the board for the establishment of PCGL's core infrastructure. PCGL will include data from multiple collections across Canada.

For Retrospective Collections, where required, additional approval or project amendments will be sought by Submitters from their local institutional REB to allow the inclusion of participants' data in PCGL, in accordance with the conditions set out in this Governance Framework (see Section 4.1). Such an approval/amendment will be the responsibility of the Submitter.

For Prospective Collections, where required, REB approval for including participants' data in PCGL should be requested at the time of the initial ethics review submission to the Submitter's local REB overseeing that collection (see Section 4.1). Additionally, participant consent forms should clearly indicate the collection's contribution to the PCGL project.

2.4. Data access and release oversight

As described in Table 1 below, data from PCGL is made available through open and controlled-access tiers, based on the type of data, its sensitivity and other requirements set by the Submitter. Data access and release under the controlled access tier is overseen by PCGL's Data Access Committee ('DAC'), as further described in Section 5.4.

The PCGL DAC has its own terms of reference detailed in Appendix A.

There may be instances where it is not always possible or equitable for the PCGL DAC to govern access to specific datasets included in PCGL. Indeed, certain collections may present with limitations (for instance due to pre-existing REB approvals, participants' consent) or with certain specifications around data governance and sovereignty (e.g. datasets from Indigenous communities, certain disease-specific collections, etc.). In these cases, the Submitter can discuss the implementation of interoperable governance oversight for data access and release with PCGL, as long as the process remains both efficient and streamlined (see Section 5.4).

2.5. Financing and Partnerships

PCGL is currently financed by the Canadian Institutes for Health Research (CIHR). Ongoing funding sources will be sought to ensure the financial sustainability of PCGL. In the future, a cost-recovery structure may be adopted, to ensure sustainability of the resource (for example, access fees may be charged to certain users or for certain uses of computing resources).

PCGL is a strongly collaborative initiative, aiming to form partnerships with national, international, public and commercial organizations to achieve its goals. Requirements for partnerships are complementary to what is covered in this Data Governance Framework. For more information on the key requirements and procedures that apply when partnering with the Pan-Canadian Genome Library, please consult PCGL's *Partnerships Framework* (link forthcoming).

3. Types of data included in the PCGL project

The following sections provide an overview of the types of data that can currently be included in the PCGL project. This list may be updated, in time.

3.1. PCGL Datasets

Data will be submitted by the Submitter to PCGL through a data submission system, which includes a Command-Line Interface (CLI) for archiving both metadata and molecular data. The submitted data encompasses study and collection-level administrative metadata, such as demographics, EDI descriptors, DAC information, clinical and phenotypical data on participants, details on biological samples, conducted genomic experiments, and data analysis information. **This data will be entered into PCGL's database as coded information, ensuring no personally identifiable data is included**. For long-term archival, the data will be stored on the SecureData4Health (SD4H) Object Storage (please see section 5.4).

The datasets included in PCGL will depend on the Submitter's local data collection practices and consent permissions (where applicable). The Submitter will use GA4GH Data Use Ontology (DUO) to declare and identify data use conditions in a standardized

way on each study. Not all data types listed in this table will be collected by every collection or for every participant.

Table 1: Overview and description of types of data available in the PCGL Library, and minimal collection data deposit requirements to be eligible for inclusion in PCGL.

(i) METADATA	Description	
a. Demographic data	Data for the categorization of the participant by means of segmenting the population (e.g. characterization by age, sex, or race).	Required
b. Clinical and phenotype data	Clinical/diagnostic data including diagnosis, treatment, phenotype, comorbidity, specimens, samples and experiments.	Required
c. Administrative data	Data about the study and program	Required
d. Experiment metadata	Information about the molecular data that is included (e.g. type of sequencer, experimental protocol)	Required
(ii) MOLECULAR DATA	**Description**	
a. Raw Genomic Data	Accepted files: FASTQ (raw sequencer output) BAM/CRAM (aligned to a reference genome) BED (for panel sequencing, describing the panel used).	Required
b. Analysis Files	Accepted files: ● GVCF / VCF files containing identified variants for an individual / collection ● Gene/transcript count files for transcriptomics data	Optional

4. Data collection and deposit in PCGL

Recruitment Sites

Participants will be recruited as part of different genomic sequencing projects (Submitters), from various sites, across Canada. The Submitter must have local research ethics approval in place for the project collecting data and contributing datasets to PCGL.

Recruitment of Potential Participants

The Submitter is responsible for the recruitment of participants for inclusion of their data in PCGL in accordance with the present Governance Framework, as well as all applicable laws, good practices, and institutional policies. PCGL will not recruit any participants directly in PCGL.

4.1. Consent Elements

4.1.1. Data from Prospective Collections

Researchers responsible for participant recruitment on behalf of the Submitter must obtain consent for the sharing of research data through PCGL. Consent forms used for prospective data collection should, at a minimum, incorporate the following core consent elements to enable broad data sharing via PCGL:

- Future health research will be conducted using participants' data through open and controlled access tiers in PCGL;

E.g.: Your data will be shared by PCGL for future research through an open access and controlled access tiers. Controlled access data will only be shared following approval by PCGL through a Data Access Committee (DAC). This DAC will verify, among other criteria, that the proposed research use is in conformity with the objectives of PCGL, your consent and that the research team applying for access has obtained the proper ethics approval. Approved researchers will sign agreements. These agreements will control how the data will be used.

- Participants' data may be shared with national and international researchers from academia, hospitals, and industry through PCGL;

E.g.: PCGL will share your data with approved researchers in Canada and around the world (which may include national and international researchers from academia, charitable organisations, hospitals, and industry.

- Data will be stored in PCGL for an indefinite period (retained until no longer useful for research purposes or as required by law);

E.g.: The data stored in PCGL will be kept indefinitely, or until it is no longer useful for research.

- Permission for the Submitter to re-contact the participant.

E.g.: I agree to be re-contacted to update my personal information, to obtain additional health information, or, to be invited to participate in new research projects.

If any of the items listed above are not included in the consent or other governance documents, datasets should not be deposited as-is in PCGL without obtaining appropriate approvals/amendments from the collection's local Research Ethics Board.
In addition to these core consent elements, Submitters of Prospective Collections are welcomed to:

- Consider agreeing that the sharing of genetic and essential metadata will be conducted through a controlled access review process spearheaded by PCGL. This does not preclude the Submitter from also making available this data through their local data access process;
- Consider signing an agreement allowing for the creation of a one-stop-shop for future data access agreements with Approved Researchers (where one host institution can sign on behalf of participating institutions).

4.1.2. ***Data from Retrospective Collections***

Given that retrospective datasets were generated before the establishment of PCGL, the consent materials initially used for them may contain different language and, in some cases, may be unclear or may not address data sharing or potential future uses.

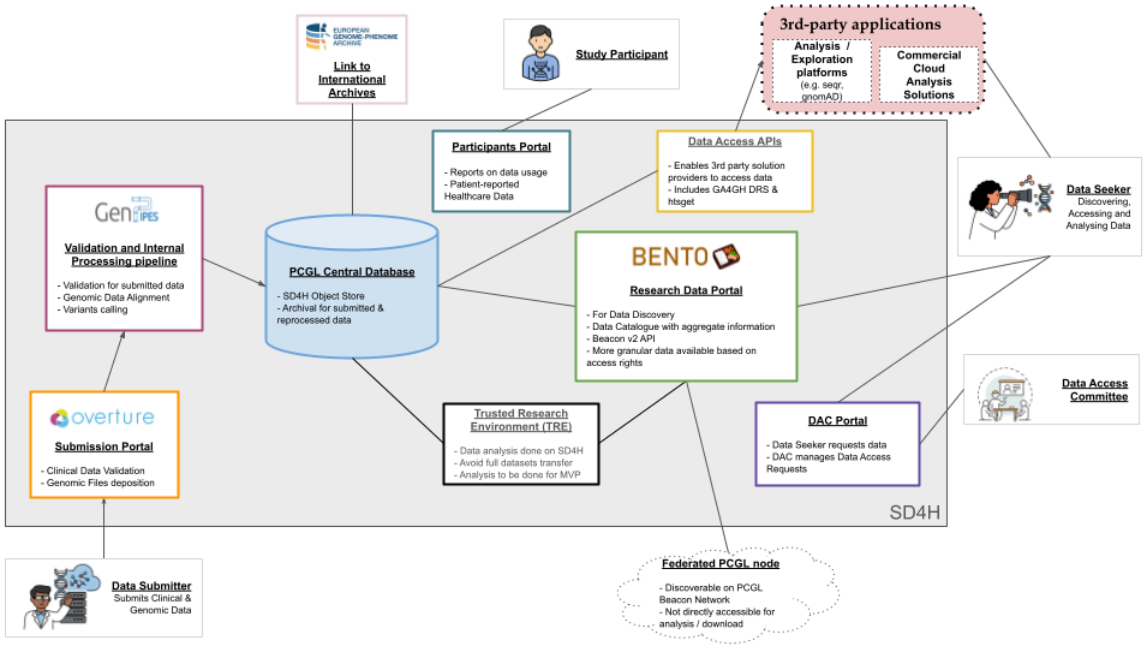
In such a case, before depositing retrospective datasets in PCGL, approval must be obtained from the Submitter's local Research Ethics Board (REB) responsible for the retrospective collection. This approval should be based on a review of the consent materials and data-sharing permissions, taking into account the core elements outlined in section 4.1.1. PCGL offers a retrospective review and inform the projects of any potential inconsistantcies.

5. Data management

5.1. Overview of data flow in PCGL

The following section summarizes the key components of the PCGL data flow, beginning with ingestion of datasets provided by Submitters, to processing, access and use. Each Submitter may have its own Standard Operating Procedure (SOP) for local data collection. However, a common data entry procedure for both integrated and federated data has been developed for contributing to PCGL, and this mechanism will be the only way to submit data to PCGL.

Figure 2: Data Flow in PCGL



5.2. Ingestion of data

Data ingestion can occur once the Submitter has collected the relevant data that is required as per Table 1 in section 3.1 and has obtained all necessary local approvals for transfer to PCGL. Two forms of ingestion exist: integration and federation.

Integrated data

Integrated data is data that can be fully transferred by the Submitter to PCGL. By entering into a data sharing agreement, McGill University will ensure custodianship over data transferred to PCGL, and data will be stored and shared in accordance with this governance framework and other PCGL policies and procedures.

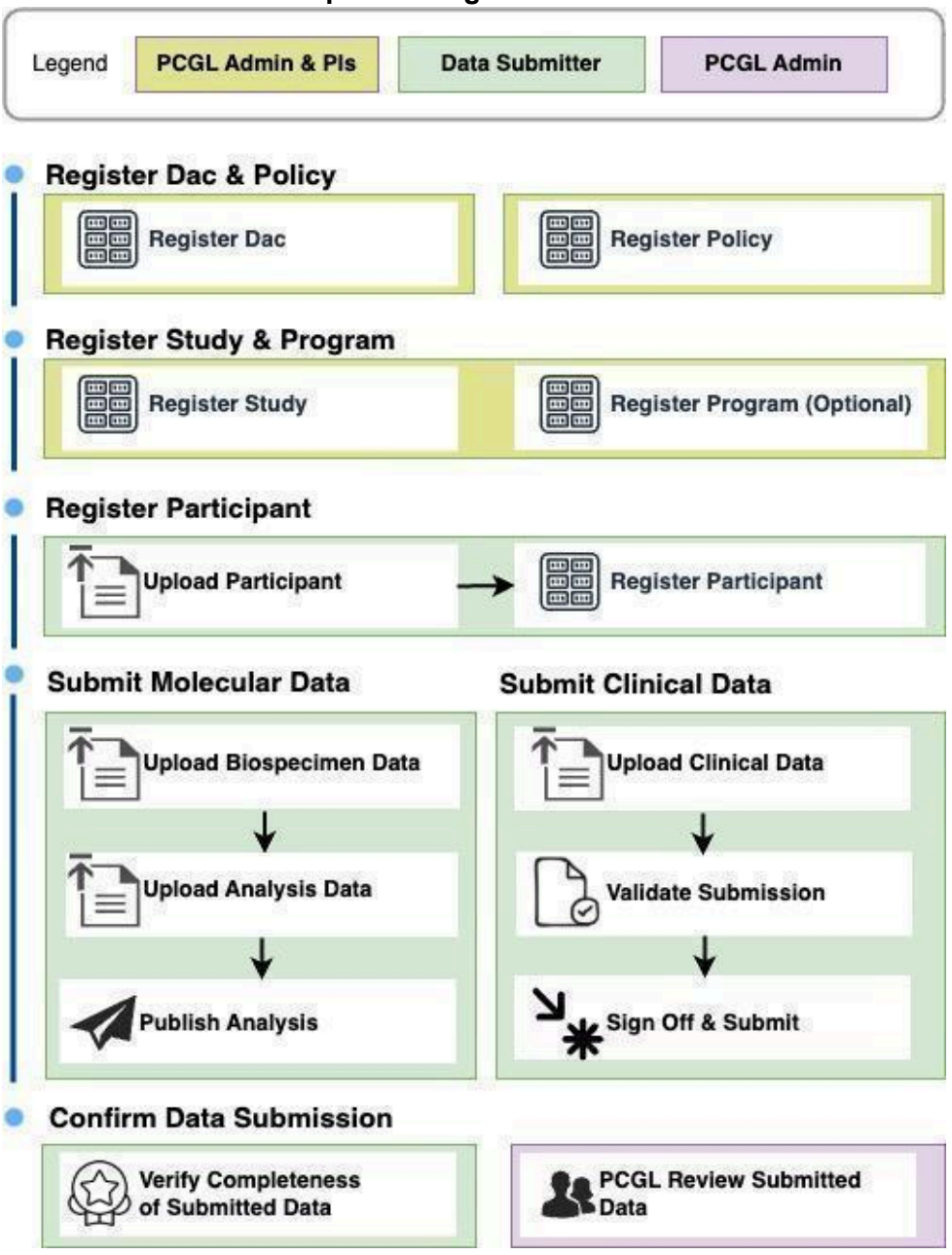
Data is ingested through data submission Command-Line Interfaces (CLI) and Application Programming Interfaces (APIs). The diagram below shows the steps of data submission. Each step is executed by a user with a specific role and associated permissions.

Data submission is split across different steps that must be executed sequentially by users with a specific role and appropriate permissions. Once proper agreements have been obtained, a PCGL administrator registers the initial information attached to the new study as a whole:

1. DAC and Policy: which platform controls data access (this can be the central PCGL DAC or an existing project-specific DAC) and what are the consented data uses (using Data Use Ontology (DUO) terms for standardized representation).
2. Study and optional program: the PCGL administrator creates a study and optionally declares what umbrella program it belongs to.

3. Once the study and policy are created, the data Submitter can enter the specific participant-level details as well as submit molecular and clinical data for each participant.
4. Data is being validated during the submission process, and the Submitter gets notified of issues. Once the data is complete and valid, the Submitter can confirm data submission, at which point the data is being sent to PCGL.

Figure 3: Data Submission Steps for Integrated Data



Federated data

Custodianship of federated datasets remains with the Submitter's institution. However, the data is available for Data Discovery within PCGL (using the PCGL Beacon Network, see 5.3).

5.3. Data storage and sharing environment

Central database
Datasets that are approved for integration within PCGL are stored in the PCGL centralized database on the SecureDataforHealth (SD4H) infrastructure.

SD4H is a cloud infrastructure distributed across several Quebec sites designed to securely analyze and share genomic and health data. It offers advanced storage solutions, high-performance computing services, and robust security measures to ensure data privacy and integrity. SD4H is hosted by and under the custodianship of McGill University (and therefore on a server located in Canada). Operated by Calcul Québec, SD4H supports researchers and healthcare professionals by providing powerful computational resources and encrypted data transfers. The infrastructure is continuously being upgraded to enhance its capabilities and meet evolving needs.

The PCGL Research Portal will be deployed on SD4H and will consist of several components that collectively provide a user-friendly environment for researchers, clinicians, and community members. These components will include: 1) an Aggregate data view, fully accessible without registration, and allowing the community to get an overview of the data available; 2) a row-level Data Exploration module for given collections a researcher has been granted access to; 3) a Data Management tool offering means to access the data; and 4) a User-centric dashboard and workspace. To enhance the accessibility and discoverability of the whole PCGL data compendium, the GA4GH Beacon Network will be implemented and open to public queries. This will also facilitate integration in other international data discovery platforms. By allowing unrestricted access to de-identified and aggregated data or publicly accessible information, the platform encourages the broader usage of PCGL data, maximizing its utility. As a key component to the federated network, the PCGL portal core architecture will be one of many endpoints facilitating access of controlled data files. To align with the broader data access mechanism, the Research Portal will interoperate with established GA4GH Authentication and Authorization Infrastructure (AAI) standards. This integration will enable users to access data which will be governed and authorized by the PCGL Data Access Committee (DAC).

5.4. Data access and use

PCGL will make accessible data through an open-access and controlled-access mechanism detailed below.

Table 2. Overview of PCGL data access tiers

TIER 1: OPEN ACCESS	
Data available under Tier 1 (Open Access)	Data that cannot be re-identified without an unreasonable effort and no particular sensitivity ("open data"), such as aggregated molecular and clinical data, including: • Participants are distributed over a list of pre-determined properties, such as disease, phenotypic condition, sex, age group. • Information on collected biological samples • Metadata on conducted genomic experiments • Aggregate variant-level data with limited phenotypic data
Who can access?	Openly accessible to anyone, through the use of APIs/PCGL website (with Terms of Use).
Access mechanism	Open access
Access Process	**Overview.** Aggregate data in PCGL will be accessible to all researchers through dedicated user interfaces and APIs. **Data Catalogue.** Researchers can explore studies through a data catalog, and request access to data. This central library facilitates the discovery and access to research materials. **Study Page.** Each study has a dedicated page displaying essential details such as a descriptive overview, provenance metadata, participant count, file types and numbers, sequencing strategies, and some limited, aggregated view on study participants, etc.
TIER 2: CONTROLLED-ACCESS	
Data available under Tier 2 (Controlled Access)	Participant-level molecular and clinical data, including: • Patient-level phenotypic, clinical and molecular data. • Molecular data will be made available in standard file formats, including GVCF/VCF files for variants, SV and CNV calls, and BAM and index files for raw reads alignments to a reference genome.
Who can access?	• Researchers approved by the PCGL DAC or collection-specific DAC (where applicable)
Access mechanism	Controlled access

Access Process	**Access to PCGL centralized datasets.** Access to participant-level detailed molecular and clinical data, such as GVCF/VCF files for variants, BAM and index files; and clinical data in standardized formats such as GA4GH Phenopackets, will be made available upon approval by the PCGL DAC. The DAC will use the corresponding GA4GH Data Use Ontology (DUO)'s codes to evaluate whether the DUO codes match between the request and the data. Access will require that a data access agreement be signed with McGill University and that the Researcher has obtained local REB approval, where applicable. The DAC terms of reference, including criteria for the review of controlled-access requests, are provided in Appendix A. **Access to PCGL federated datasets.** Datasets that are discovered on the PCGL Research Portal through a Beacon Network query, and that are located outside of PCGL's central node on SD4H, will be accessible by placing a data access request directly to the custodians of those datasets. The PCGL Research Portal will provide a link to the proper online resource to place such a data access request, as the PCGL DAC will not review them.

5.5. Data discovery on other databases

In order to facilitate international data sharing and create shared resources for data discovery, efforts will be made to integrate datasets with other Canadian and international databases, where possible (e.g. European Genome Archive). This will be facilitated by implementing programmatic interfaces that are compatible with these other initiatives' platforms, such as the GA4GH Beacon API.

6. Database closure

Data integrated into PCGL is currently under the custodianship of McGill University.

If PCGL discontinues its activities at any point in the future, efforts will be made to transfer the data to a Canadian entity that agrees to follow PCGL's mission of promoting Canadian genomic research through a common genomic data platform, to its policies and to the terms of participants' consent. Before any transfer of data custodianship or responsibilities for maintaining PCGL takes place, each Submitter will be notified and

given the possibility to withdraw their data from the platform. Any decision regarding the transfer or closure of the database will involve McGill University, as the current custodian of PCGL data and infrastructure. For the long-term sustainability of PCGL, it is also possible that a new entity may become the data custodian. In that case, PCGL will also notify all Submitters.

7. Protection of participant privacy and data security measures

To further protect participants' information, upon inclusion of collection data in PCGL, data will be coded according to the Submitter's procedure. If the Submitter's REB or institution requires data to be double coded prior to depositing data in the PCGL, double coding should be done by the Submitter, which will ultimately be responsible for providing this new, second code to PCGL.

The only link between the participant's identifying information and the coded data stored in the PCGL is the participant code (participant ID). The coded data is thereafter imported and stored in PCGL's database. The key to link the participant's PCGL number and collection participant ID (e.g. linking log) is held locally by the Submitter's principal investigator. PCGL will not collect or store information that directly identifies participants (e.g. names, contact information, healthcare numbers, etc.). Therefore, the possibility of obtaining nominal information on participants by consulting the PCGL database is very low.

Only the authorized members of PCGL's team and Researchers approved by the PCGL DAC will have access to authorized sets of coded clinical/phenotypic and molecular data. Members of PCGL's team are only authorized to use and access the data to the extent that it is necessary to realize their administrative function within PCGL. This access is conditioned to the prior signature of a *PCGL confidentiality agreement*. To use the data for research purposes, they need to follow the standard procedure established by this framework.

Moreover, McGill University is committed to performing a Privacy Impact Assessment (PIA) as per the provincial law. The PIA will consider all types of factors that can have an impact on the privacy of the participants, especially through risks analysis and mitigation.

8. Potential benefits and risks to research participants

8.1. Benefits

There may be no direct benefits to the participants contributing data to PCGL.

PCGL will facilitate data sharing of genomic data in Canada and world-wide, enabling the discovery of molecular-based treatments and outcomes, by building the

infrastructure and tools needed to improve research on genomics-based health conditions.

8.2. Physical risks

PCGL is not responsible for biospecimen sampling procedures, sequencing and clinical data collection. Sampling procedures and risks associated with such sampling are detailed, as relevant, as part of each Submitter's protocol and consent forms.

There is no physical risk involved in taking part in PCGL.

8.3. Informational risks

Much like fingerprints, it is possible to identify someone if certain data about that person are put together from different sources. While strict data security and privacy measures are in place by PCGL to protect the participants' privacy, there is always a small risk that the participants' data may lead to their re-identification. Moreover, as technology advances, there may be new ways of re-identifying participants that cannot be foreseen today. In case of a breach of confidentiality, personal health information may affect the participant's insurability or employment. However, Canada has enacted legislation that provides against genetic discrimination based on the result of genetic testing, mitigating this risk.

There is a possibility that researchers accessing PCGL will be able to link individual-level datasets to a participant's identity based on molecular and phenotypic data in PCGL. However, upon accessing controlled-access data from PCGL, researchers will have to sign agreements that prevent them from intentionally attempting to do so.

If a privacy breach occurs, PCGL will contact the Submitter within reasonable timelines to inform it of the breach and explain what data may have been compromised. It is the responsibility of the Submitter to communicate this information to participants as only they can recontact the participant.

9. Duration of conservation of data

Data stored in PCGL will be kept indefinitely or until no longer useful for research purposes. However, data can be removed from PCGL if the principal investigator of a Submitter informs PCGL that a participant has withdrawn consent, that the collection's institution or REB has requested that data be removed, or in the event that PCGL stops its activities (see Section 6). Limitations on data withdrawal are outlined in the next section.

10. Management of participant withdrawal

Participants have the right to withdraw data from PCGL at any time and without providing any reason. Participants can withdraw their data from PCGL by informing the Submitter who is contributing to PCGL. The Submitter will then contact PCGL to request withdrawal and that data be destroyed and removed from the Library. However, data that has already been accessed for research cannot be removed or destroyed to preserve the scientific integrity of the analysis.

11. Results and intellectual property

Intellectual Property rights may not be claimed on the data stored in PCGL. Intellectual Property rights on derived data is acceptable, if it does not impede data usage by the researchers accessing PCGL.

Patent protection will not be sought for any of the gene discoveries developed as a result of the use of PCGL data. PCGL believes that open science has the most rapid impact for patients. This does not mean that eventual products, drugs or tests may not be commercialized by Approved Researchers.

12. Useful References

- PCGL Partnership's Framework
- Global Alliance for Genomics and Health, Framework for responsible sharing of genomic and health-related data (2014)
- CMAJ Article - Core elements of participant consent documents for Canadian human genomics research and the National Human Genome Library: guidance for policy
- FAIR Principles
- GA4GH Consent Policy
- GA4GH Diversity in Datasets
- GA4GH's Data Access Committee Guiding Principles and Procedural Standards Policy

Appendix A: PCGL Data Access Committee Terms of Reference

1. Role
The role of the PCGL Data Access Committee (DAC) is to receive access requests for projects proposing to use controlled-access data stored centrally by PCGL (McGill custodianship). The DAC assesses the feasibility of the proposal, reviews and approves, conditionally approves or denies the request, according to criteria set out in section 7. Moving forward, the Data Access Committee will fully implement the standards set out in GA4GH's Data Access Committee *Guiding Principles and Procedural Standards Policy* (DACReS).

2. Term
These Terms of Reference are effective from [xxxxx].

3. Membership
The Committee members (including the Chair) will be appointed by PCGL's Executive Committee. The Committee shall be composed of a minimum of five (5) members, in majority independent from PCGL, including:

- At least one (1) genomic researchers [voting member];
- At least one (1) expert in the technical infrastructure of the PCGL [voting member];
- At least one (1) expert on the legal/ethical aspect of genomic research, data sharing, privacy and data protection [voting member];
- One (1) investigator representing the PCGL program [voting member];
- At least one (1) patient/participant representative or member of the public or communities of interest [voting member].

An Access Administrator will facilitate the work of the Data Access Committee.

For certain access applications, the DAC Chair may call upon outside experts. Such an expert will not be a voting member on the Committee, but invited to provide background expertise required to review the application.

Committee members (including the DAC Chair) will serve a two-year term, renewable by the Executive Committee.

4. Responsibilities
The membership of the DAC commits to:

- Attending all scheduled DAC meetings or, in case of absence, providing detailed notes and decisions to the DAC ahead of the meeting;
- Sharing all communications and information across all DAC members;
- Making timely decisions regarding data access requests.

Members of the DAC expect:

- That each DAC member is provided with complete, accurate and meaningful information in a timely manner prior to each meeting;
- To be given reasonable time to make key decisions.

5. Meetings

A meeting quorum is of 3 members of the DAC. Meetings are held monthly via videoconference. Decisions are made by consensus (i.e. members are satisfied with the decision even though it may not be their first choice). If consensus is not possible, a vote will take place. In case of a tie, the DAC Chair makes the final decision.

6. Users involved in the application

The following are a list of the users involved in an application and their requirements:

- Applicant/Principal Investigator: An independent researcher who is affiliated with a legal entity (e.g. university professor, researcher in a private company, independent researchers able to apply for federal research grants, etc.)
- Institutional Representative: A qualified representative of a legal entity who has the administrative power to legally commit that entity to the terms and conditions of the Data Access Agreement (e.g. Vice-President Research, a Research Director, or a Contracts Officer for the entity)
- Collaborators: All investigators, collaborators, research staff at the Applicant's institute (including post-docs) and students (including graduate students), who will have access to PCGL's Controlled Data in order to work on the research project

7. DAC Approval Process

The DAC Chair will receive an email notifying them that the DAC application has been submitted. During the review period, the online application will be locked and uneditable.

The following steps summarize the controlled-access application process:

1. The Principle Investigator (PI) must submit a completed Application Form requesting access to controlled-access data to the DAC using the PCGL Portal. A new application form must be submitted for each new study. As part of the application, the PI should list all other collaborators who require access to the data (ex: researchers, employees, laboratory personnel, etc.).
2. As part of its review, the DAC may consider the following information:
   - Compatibility of the proposal with the objectives of the PCGL;
   - Justification that the project requires access to PCGL datasets;
   - Qualifications of the PI to undertake the proposed project;

- Number and names of research team members that will have access to data and rationale for their access;
- Scientific merit and feasibility of the proposed project;
- Proof of local research ethics committee approval for the proposed project (or justification as to why such approval is not required); and

3. Upon approval by the DAC, a Data Access Agreement is entered into between the Researcher and McGill.
4. Controlled access to PCGL data is granted for a period of 1 year (which may be renewed, upon request).
5. Approved Researchers may contact the DAC for any amendment to the initial application.

8. Amendment, Modification or Variation

These Terms of Reference may be amended, varied or modified in writing after consultation and agreement by the DAC and the Oversight Board. A public notice highlighting the amendments will be published on the Access Portal.